

A bearing fault diagnosis method for high-noise and unbalanced dataset

Rui Wang

Security Division, CRRC Dalian Institute Co., Ltd., Dalian, China

Shunjie Zhang

*State Key Laboratory of Rail Traffic Control and Safety,
Beijing Jiaotong University, Beijing, China*

Shengqiang Liu and Weidong Liu

Security Division, CRRC Dalian Institute Co., Ltd., Dalian, China, and

Ao Ding

*State Key Laboratory of Rail Traffic Control and Safety,
Beijing Jiaotong University, Beijing, China*

Abstract

Purpose – The purpose is using generative adversarial network (GAN) to solve the problem of sample augmentation in the case of imbalanced bearing fault data sets and improving residual network is used to improve the diagnostic accuracy of the bearing fault intelligent diagnosis model in the environment of high signal noise.

Design/methodology/approach – A bearing vibration data generation model based on conditional GAN (CGAN) framework is proposed. The method generates data based on the adversarial mechanism of GANs and uses a small number of real samples to generate data, thereby effectively expanding imbalanced data sets. Combined with the data augmentation method based on CGAN, a fault diagnosis model of rolling bearing under the condition of data imbalance based on CGAN and improved residual network with attention mechanism is proposed.

Findings – The method proposed in this paper is verified by the western reserve data set and the truck bearing test bench data set, proving that the CGAN-based data generation method can form a high-quality augmented data set, while the CGAN-based and improved residual with attention mechanism. The diagnostic model of the network has better diagnostic accuracy under low signal-to-noise ratio samples.

Originality/value – A bearing vibration data generation model based on CGAN framework is proposed. The method generates data based on the adversarial mechanism of GAN and uses a small number of real samples to generate data, thereby effectively expanding imbalanced data sets. Combined with the data augmentation method based on CGAN, a fault diagnosis model of rolling bearing under the condition of data imbalance based on CGAN and improved residual network with attention mechanism is proposed.

Keywords Imbalanced samples, CGAN, Channel attention mechanism, Residual network, Bearing fault diagnosis

Paper type Research paper



1. Introduction

Railway transportation has the advantages of strong carrying capacity, fast transportation speed, high safety, punctuality, high efficiency and environmental protection. Therefore, rail transit plays an irreplaceable role in many countries. Safety is the premise of rail transportation and the core competitiveness of the rail system. However, factors such as high-intensity and -density use and changing environments will inevitably reduce the safety of railway train operations. Train wheelset bearings are one of the key components of trains. If the bearing fails, it will have very serious consequences. Regularly dismantling, inspecting and reassembling the bearings requires a lot of work. In addition, this method not only requires the maintenance personnel to have rich experience and professional knowledge but also has the possibility of missing bearing faults between two inspection intervals, resulting in operational accidents. Therefore, it is necessary to study the fault diagnosis technology of real-time monitoring of train wheelset bearings.

With the development of deep learning technology, data-driven fault diagnosis technology has attracted more and more researchers' attention because it does not require artificial design features and has good generalization ability. [Zhao et al. \(2018\)](#) developed dynamically weighted wavelet coefficients to improve the performance of ResNet-based diagnostic models and obtained higher fault diagnosis accuracy for planetary gearboxes in severe noise environments compared to other deep learning-based methods. [Qiao et al. \(2019\)](#) proposed an adaptive weighted multiscale convolutional neural network for more accurate fault diagnosis under complex working conditions. [Udmale et al. \(2019\)](#) realized intelligent recognition of bearing faults based on spectral kurtosis and convolutional neural network. [Chen et al. \(2019\)](#) used continuous wavelet transform to preprocess the data, then used a convolutional network to extract features and finally used an extreme learning machine as a strong classifier to achieve high-accuracy bearing fault diagnosis. [Liu et al. \(2020\)](#) integrated the two problems of fault diagnosis and remaining useful life prediction and designed a joint loss convolutional neural network, which effectively improved the performance of the convolutional model for the two tasks. [Xiong et al. \(2020\)](#) implemented the wavelet packet transform in the form of convolution and embedded it in the convolutional neural network, which improved the performance of model diagnosis while ensuring end-to-end fault diagnosis. [You et al. \(2020\)](#) improved the activation function of the fault diagnosis model based on convolutional neural network, making the gradient easier to propagate, and achieved excellent results. [Kou et al. \(2020\)](#) used a convolutional neural network for multisensor data fusion and feature extraction and realized fault diagnosis of the train bogie rotating mechanism.

However, the application of these algorithms to practical engineering still faces the following challenges:

- In the actual train operation, the fault samples are far less than the normal samples, resulting in data imbalance. At present, most of the fault diagnosis models in research need to be trained with balanced data sets to achieve better results, and the impact of unbalanced data sets on fault diagnosis based on deep learning has not been fully considered.
- In the current fault diagnosis extraction research of deep learning method, the vibration signal feature extraction work needs to consider how to further extract high-dimensional features under large noise conditions to meet the diagnostic requirements.
- To solve the above problems, the following research is carried out in this paper.
- A bearing vibration data generation model based on the conditional generative adversarial network (CGAN) framework is proposed. The method generates data

based on the adversarial mechanism of generative adversarial networks (GAN) and uses a small number of real samples to generate data, thereby effectively expanding imbalanced data sets.

- Combined with the data augmentation method based on CGAN, a fault diagnosis model of rolling bearing under the condition of data imbalance based on CGAN and improved residual network with channel attention mechanism is proposed.

The method proposed in this paper is verified by two data sets, proving that the CGAN-based data generation method can form a high-quality augmented data set. At the same time, the diagnostic model based on CGAN and the improved residual network with channel attention mechanism has excellent diagnostic accuracy under low signal-to-noise ratio.

2. Theoretical basis

2.1 Data imbalance problems and solutions

In practical engineering, bearings are in normal operation most of the time. Data collection for failed bearings is often difficult. It leads to an imbalance in the number of normal samples and fault samples in the actual engineering data set. This imbalance in the data may lead to poor diagnostic performance of the model. Therefore, more and more researchers have begun to focus their research on bearing fault diagnosis under data imbalance. At this stage, the data imbalance problem is mostly solved from the data and algorithm levels.

At the data level, various data augmentation techniques are used on the original data set to augment the sample size of quantitatively disadvantaged classes and, finally, achieve the balance of various classes of samples in the model training data set. Such methods include synthetic minority over-sampling technique (Raghuwanshi and Shukla, 2020), adaptive synthetic sampling (Gonzalez *et al.*, 2019), etc. However, the synthetic minority over-sampling technique is prone to cause the problem that the generated samples overlap with the original samples. And adaptive synthetic sampling is susceptible to outlier interference. The available information supplemented by these traditional methods is very limited and does not contribute significantly to the improvement of fault diagnosis accuracy.

At the algorithm level, in the case of imbalance, the improvement of the algorithm is mainly based on ensemble classification and sample weighting. For instance, Chen *et al.* (2021) proposed a weighted balanced distribution adaptation method (MC-W-BDA), in which enough base classifiers are obtained by random sampling with different training sample sets trained in the replicated kernel Hilbert space. Multiclassifier ensemble, by integrating appropriate base classifiers into strong classifiers through a multiclassifier ensemble strategy, to solve the data imbalance problem. Lin *et al.* (2020) proposed a novel loss function to make the deep model pay more attention to the classes with a small number of samples during the training process.

2.2 Conditional generative adversarial network

The GAN is a network architecture based on the idea of a zero-sum game (Goodfellow *et al.*, 2014). As an unsupervised data generation model, GAN does not rely on any prior assumptions and can generate high-quality new data that conforms to the sample distribution, attracting attention from both academia and engineering. Its basic frame diagram is shown in Figure 1. The model mainly includes two parts: generator (G) and discriminator (D). The random noise z generates samples after passing through G , and the discriminator D recognizes real samples and generated samples as much as possible. In this process, the purpose of G and D are opposed to each other. By continuously training separately and alternately, G and D optimize their parameters in the process of playing against each other, and finally, G and D are in nash equilibrium.

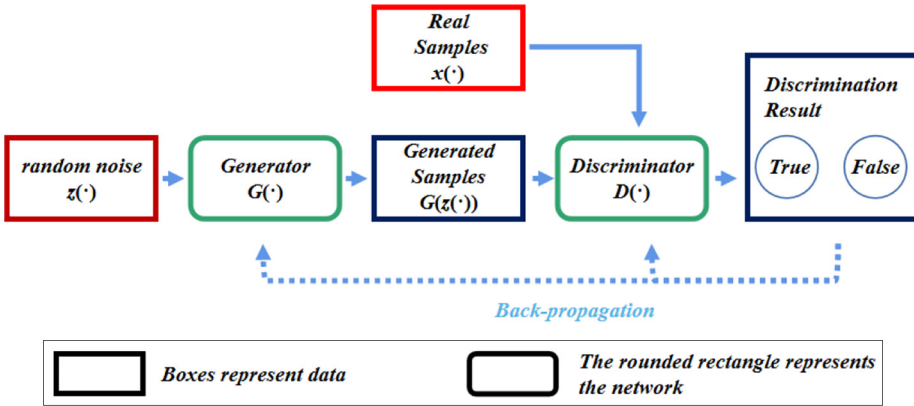


Figure 1. Framework of GAN

CGAN is an improvement of GAN (Ahsan *et al.*, 2022). By introducing label information into the GAN's generator and discriminator, the generated results can be controlled by adjusting the labels as samples are generated. The network architecture diagram of the CGAN is shown in Figure 2. The objective function of CGAN can be expressed in the form of equation 1, where G and D represent the conditional probability obtained after adding additional conditional information. In this paper, CGAN is introduced to solve the problem of data imbalance because CGAN can directionally expand a certain minority class data while avoiding the loss of computational cost caused by training multiple GANs:

$$\arg \min_{\theta_G} \max_{\theta_D} L_D(\theta_G, \theta_D) = \frac{1}{n} \sum_{k=1}^n \log D(x|y; \theta_D) + \frac{1}{m} \sum_{i=1}^m \log (1 - D(G(z(\cdot)|y; \theta_G)|y; \theta_D)) \quad (1)$$

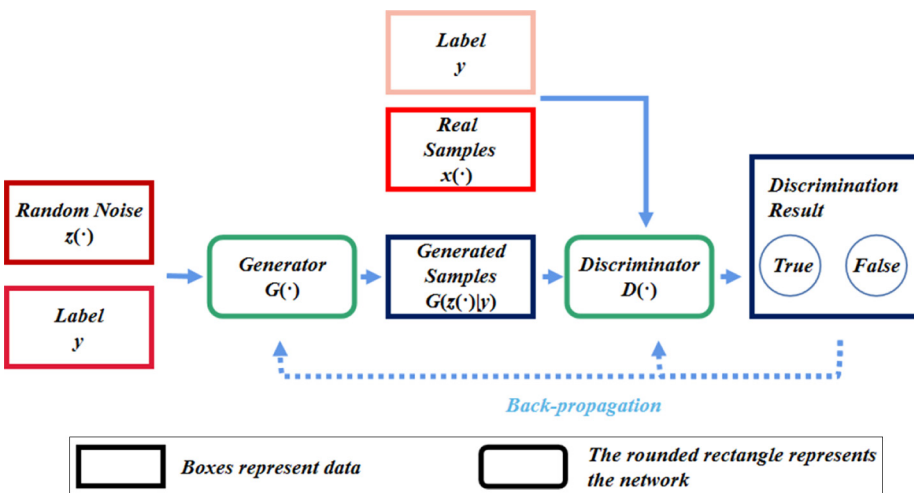


Figure 2. Framework of CGAN

2.3 Residual network

Deep residual network solves the problems of gradient dispersion, gradient explosion or network degradation in neural networks by introducing shortcut connections (He *et al.*, 2016). The original input information, or the output of a certain layer, can be directly passed to the bottom layer of the network through identity mapping, which makes the training of deep networks easier and more stable. The residual block is the basic structure of the residual network, as shown in Figure 3. A deeper model means a stronger nonlinear expression ability of the model, which helps to improve the feature extraction and recognition ability of the convolutional model. Therefore, the diagnostic model proposed in this paper introduces the residual structure.

2.4 Channel attention mechanism

By simulating the characteristics of the human brain, the attention mechanism makes the model pay more attention to the features that contribute more to the task and ignore the features that contribute less to the task (Niu *et al.*, 2021). Among them, squeeze-and-excitation networks (SENet) are currently widely used in various convolutional neural networks (Hu *et al.*, 2020). The basic unit of SENet is the SE Block, and its structure is shown in Figure 4. SE Block mainly includes three parts: squeeze, excitation and reweight. In the squeeze part, feature compression is performed within the channel. In the excitation part, calculate the importance of each channel. In the reweight part, the implementation weights the channels based on importance. Considering the importance of each channel helps to improve the recognition ability of the recognition model. Therefore, in the diagnostic model proposed in this paper, the SE Block is introduced.

3. Solution to the data imbalance problem based on conditional generative adversarial network

In this paper, a CGAN-based model for generating vibration signals of train wheelset bearings in the time domain is proposed. The model trains the generator and discriminator of CGAN by inputting the time-domain vibration data under different fault states of the rolling bearing and its corresponding fault type labels. The trained generator can generate high quality time-domain vibration signals by inputting fault state labels to effectively

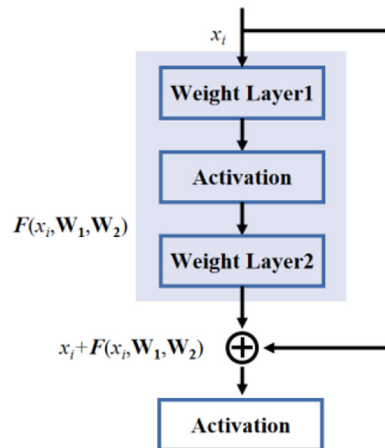


Figure 3.
Framework of the
residual block

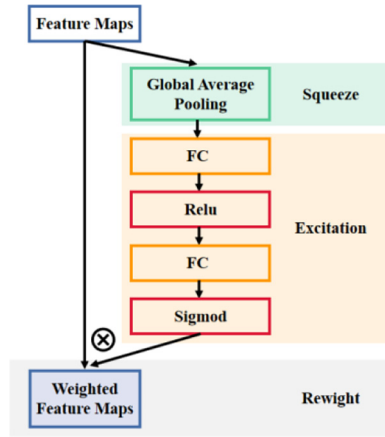


Figure 4. Framework of the SE block

expand data samples and achieve the purpose of equalizing the number of samples. The structure of the model is shown in Figure 5. The loss function of the model is shown in equation 1, and the optimizer can use Adam.

The generator of CGAN consists of four fully connected layers. The first three layers are activated using the Leaky Relu function, and the batch normalization layer is introduced to accelerate the model convergence. The last layer uses linear activation. The input to the generator is a $1 \times n$ random noise, where n is usually a large positive integer. And the output of the generator is a $1 \times m$ signal, where m is the length of the signal to be generated that corresponds to the true sample.

The discriminator of CGAN is also composed of four fully connected layers, using the Leaky Relu activation function and introducing the dropout layer to alleviate overfitting. The discriminator is connected to the output layer, which can also be regarded as a fully connected layer, containing two neurons and activated using the softmax function. The input to the discriminator is a $1 \times m$ sample. The input samples include real samples and samples generated by the generator. The output layer outputs a vector containing two

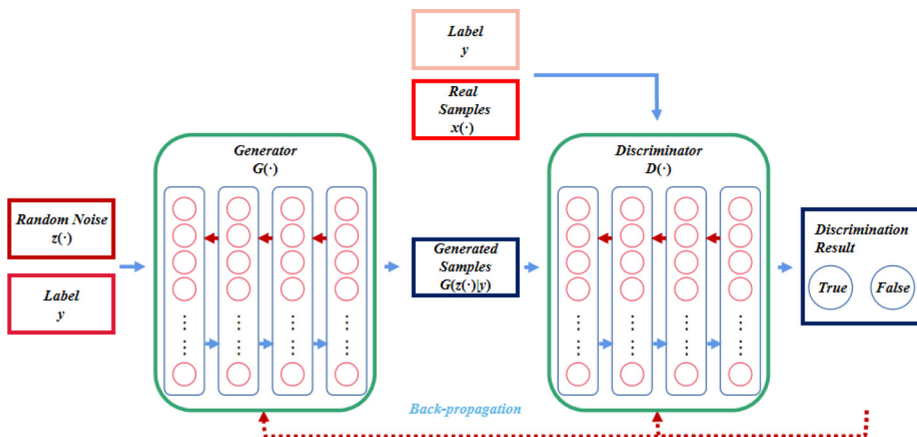


Figure 5. Model structure of CGAN

elements, and the elements in the vector represent the probability that the discriminator considers the current sample to be the real sample and the generated sample.

4. Bearing fault diagnosis model based on fusion residual network and SE blocks

The vibration signal of real train wheelset bearing often has a much larger noise than the experimental data. And when using CGAN for data augmentation, the noise is not filtered out. Because of this situation, the fault diagnosis model of train wheelset bearings may require better signal feature extraction capability. When using the residual network to build an end-to-end fault diagnosis model, by deepening the network structure, high-dimensional features with richer fault information can be extracted from the input signal, and there is no need to worry about the gradient problem. The SE block is a simple and effective channel attention mechanism block, which has the advantage of low complexity. It can realize adaptive weighting of the feature maps of each channel and make the model pay more attention to the key features that affect the accuracy of fault diagnosis. Based on the above ideas, this paper proposes a novel bearing fault diagnosis model, which combines the residual network and SE attention mechanism structure. Its specific structure is shown in [Figure 6](#).

The model uses the raw vibration data folded into two dimensions as model input (such as the vibration signals of 100 sample points reshaped into a 10×10 matrix). Fault features are efficiently extracted by stacking 2D convolutional layers activated by the Relu function. And the batch normalization layers are introduced to speed up the training process of deep learning to avoid overfitting. Pooling layers reduce the number of parameters to speed up the training process. Dropout layers avoid overfitting. And two fully connected layer constructs a classifier. Among them, the last fully connected layer is activated by the softmax function, and the number of neurons in this layer is determined by the number of classes of samples in the recognition task. In the model, residual blocks and SE blocks are used alternately. On the one hand, the gradient problem in model training is effectively avoided, and on the other hand, valuable features are enhanced by weighting the features to achieve better diagnostic performance. The diagnostic model can use multiclass cross-entropy as the loss function and be trained using the Adam optimizer. The fault diagnosis model can be used in conjunction with the CGAN data generation model to solve the bearing fault diagnosis in the case of data imbalance and low signal-to-noise ratio of data.

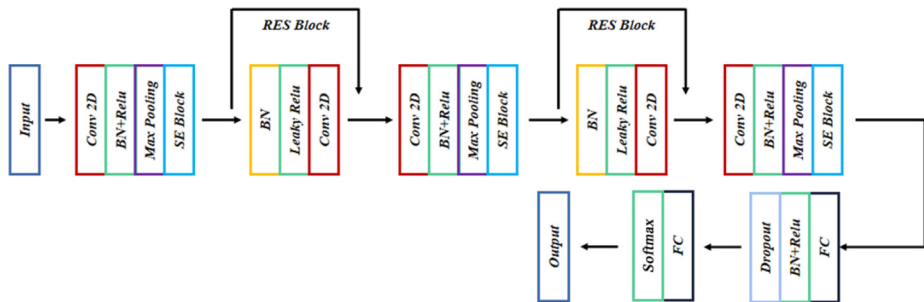


Figure 6.
Model structure of the fault diagnosis model

5. Experiments and results

5.1 Experimental data set

To verify the quality of the data generated by the CGAN model and the performance of the improved residual network model with SE attention mechanism in the task of bearing fault diagnosis in a high-noise environment under the data augmentation of the CGAN model, experiments are carried out. In the experiment, the current mainstream bearing vibration data set of Case Western Reserve University (CWRU) and a vibration signal data set collected from an actual train fault bearing (QDSF) were used.

Case Western Reserve University data set: This data set is an industry-recognized standard data set for bearing fault detection. It contains a large number of rolling bearing vibration signals in normal and fault conditions. In recent years, a large number of data-driven bearing fault diagnosis algorithms have used this data set for algorithm training and validation (Neupane *et al.*, 2021; Zhang *et al.*, 2021; Liu *et al.*, 2022). Deep groove ball bearings use electron discharge machining single point damage to artificially form faults. The data in the data set is divided into four states: normal, inner race fault, outer race fault and ball fault. There are three diameters of artificial faults, 0.007, 0.014 and 0.021 inches to represent minor, medium and serious faults, respectively. So, data can be labeled with ten kinds of labels. And the ten labels can be noted as: “0.007-IRF,” “0.007-ORF,” “0.007-BF,” “0.014-IRF,” “0.014-ORF,” “0.014-BF,” “0.021-IRF,” “0.021-ORF,” “0.021-BF” and “Normal.” The data acquisition process includes four load conditions (respectively, corresponding to 1,797, 1,772, 1,750 and 1,730 rpm), simulating real changing conditions. The acceleration vibration signal comes from the drive end and the fan end. This paper uses the drive end data, and the sampling frequency is 12 KHz.

QDSF data set: The data in this data set is from vibration signals collected from the Qingdao Sifang freight train bearing test platform. The faulty bearings used in the data set collection are collected from the actual freight trains. The bearings are SKF197726 type double row tapered roller bearings. The data set includes different degrees of ball faulty, inner race faulty, outer race faulty, cage faulty and normal bearing vibration signals. The photos of the four faults are shown in Figure 7. According to the width of the fault, the rolling element fault has two degrees of 0.001 and 0.0105 mm, the inner ring fault has two degrees of 0.135 and 0.45 mm and the outer ring fault has two degrees of 0.3 and 0.45 mm. Therefore, there are eight kinds of data with different labels, denoted as: “0.001-BF,” “0.0105-BF,” “0.135-IRF,” “0.45-IRF,” “0.3-ORF,” “0.45-ORF,” “CF” and “normal.” The vibration signal was collected using the experiment platform shown in Figure 8. The experimental platform is mainly composed of the drive motor, transmission device, axle, support bearing, axial force loading device, radial force loading device, experimental axle box assembly, cooling fan and other parts. During the experiment, the bearing rotation speed was set to three levels, which were 90 (589 rpm), 120 (786 rpm) and 150 km/h (983 rpm). Based on the vertical loads applied by the train under no-load, half-load, full-load and 20% overload conditions, the corresponding vertical loads under these conditions are 56, 146, 236 and 272 kN, respectively. To simulate the turning state of the train, the lateral load is set as a cycle from 0 to 20 to 0 kN. The sampling frequency of the signal is 12.8 KHz. Compared with the CWRU data set, this data set is closer to the real situation and contains more noise.

5.2 Quality of the vibration signal generated by the conditional generative adversarial network model

In this experiment, a CGAN model is first constructed. The input of the generative model is 128-dimensional Gaussian noise and fault class label information. The output is a one-dimensional generated vibration signal of length 10,000 that conforms to the label. The

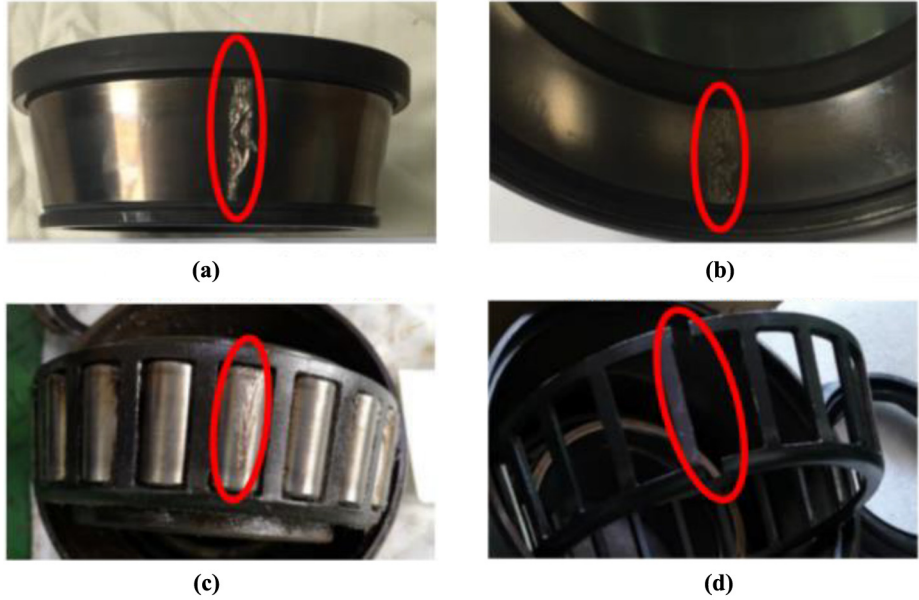


Figure 7.
Details of bearing
failures

Notes: (a) Inner race stripping fault; (b) outer race stripping fault; (c) rolling element stripping fault; (d) cage fracture fault

Figure 8.
Photograph of the
experiment platform



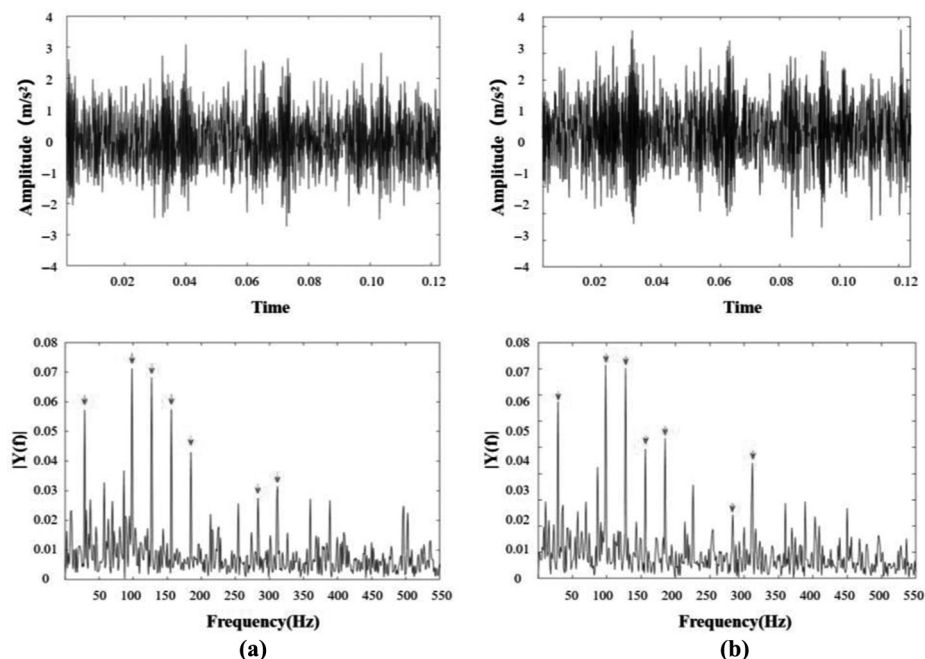
input of the discriminant model is the pseudo vibration signal generated by the generator with label information and the real vibration signal of the same label. These signals all have a length of 10,000. When the number of iterations reaches 3,000 and the loss function value of CGAN is less than the convergence threshold of 0.5, the model training is considered complete. By comparing the squared envelope analysis spectrum of the generated signal and the real signal, the quality of the signal generated by the CGAN model can be evaluated (Antoni, 2007).

For the CWRU data set, 1,000 samples of each fault class were randomly intercepted in the experiment to train CGAN. Figures 9–11 show time-domain plots and spectrograms from squared envelope analysis of the CGAN-generated signals and real signals (labeled “0.007-IRF,” “0.007-ORF,” “0.007-BF,” respectively). It is not difficult to see that the amplitude of the pseudo vibration signal generated by the CGAN model is slightly different from the real signal in the time domain, but the general characteristics of the signal are consistent. Besides, in the frequency domain, it can be seen that the CGAN-generated signal has peaks similar to the real signal at the fault characteristic frequency and the multiplier frequency.

For the QDSF data set, 1,000 samples of each fault class are randomly intercepted in the experiments to train the CGAN. Figures 12–14 show the time domain and spectrograms from squared envelope analysis of the real signals and CGAN-generated signals (labeled “0.45-IRF,” “0.30-ORF,” “0.105-BF,” respectively). The conclusions drawn from the results of this group of experiments are consistent with the CWRU data set. Observed in the time and frequency domains, it can be seen that the quality of the generated signal is excellent. It should be noted that, in terms of the consistency of fault feature frequencies, CGAN’s ability to generate the QDSF data set is worse than that of the CWRU data set. It is because the QDSF data is closer to the ground truth and contains a lot of noise, making the faulty features of the CGAN-generated signals less obvious.

5.3 Performance of fault diagnosis model

In this experiment, considering the characteristics of the data, under the guidance of experience and hyperparameter fine-tuning results, a bearing fault diagnosis model based



Notes: (a) Signal generated by CGAN; (b) real signal

Figure 9.
Time domain and
frequency domain
waveforms of “0.007-
ball” type vibration
signals generated by
CGAN in Case
Western Reserve
University bearing
data set

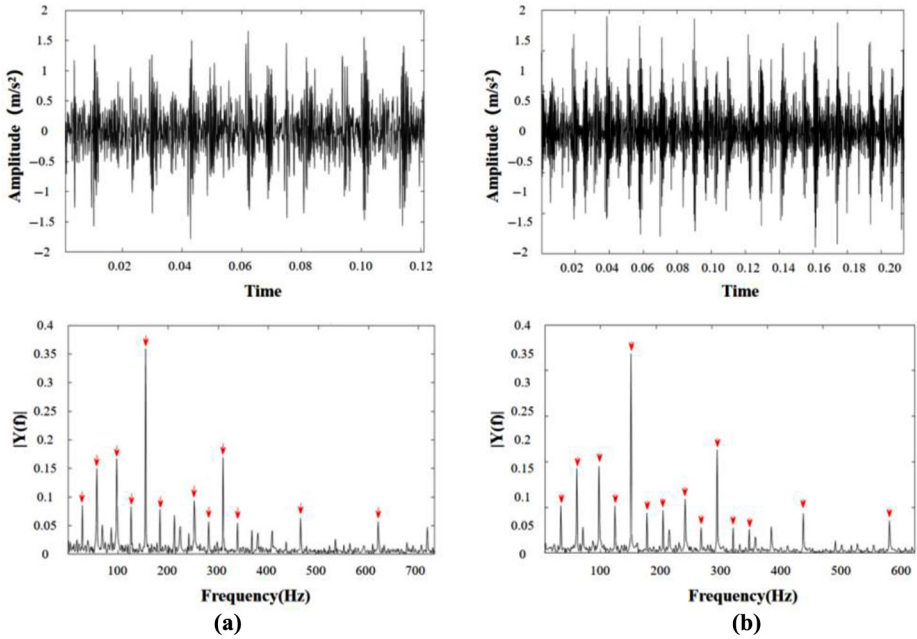


Figure 10.
Time domain and frequency domain waveforms of “0.007-inner” type vibration signals generated by CGAN in Case Western Reserve University bearing data set

Notes: (a) Signal generated by CGAN; (b) real signal

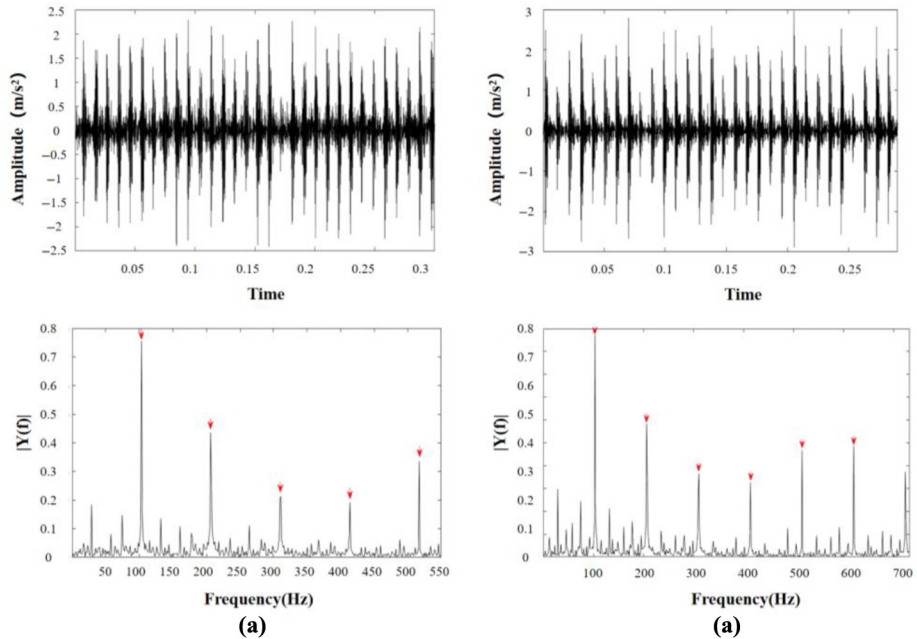


Figure 11.
Time domain and frequency domain waveforms of “0.007-outer” type vibration signals generated by CGAN in Case Western Reserve University bearing data set

Notes: (a) Signal generated by CGAN; (b) real signal

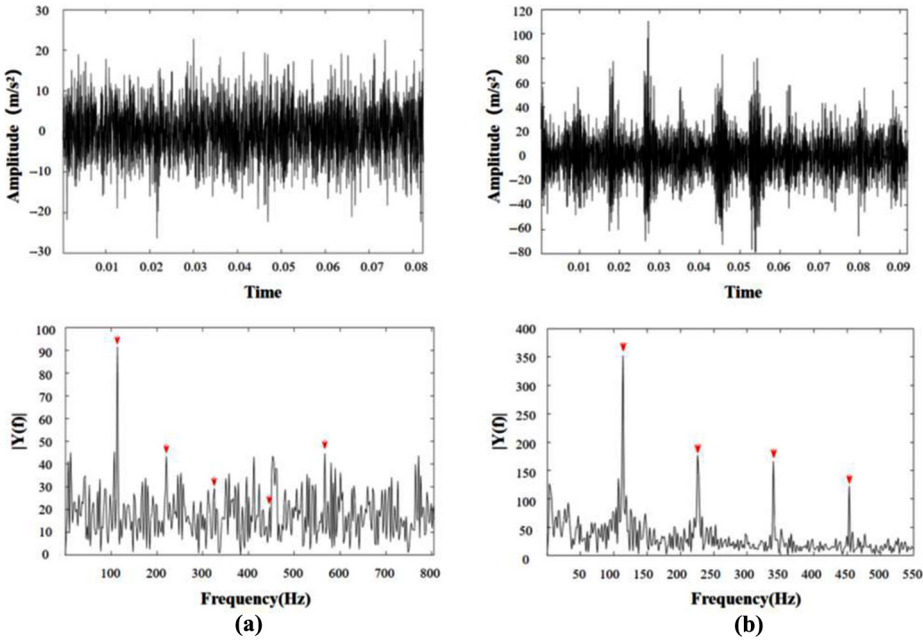


Figure 12. Time domain and frequency domain waveforms of “0.30-outer” type vibration signals in real railway wheelset bearing database

Notes: (a) Signal generated by CGAN; (b) real signal

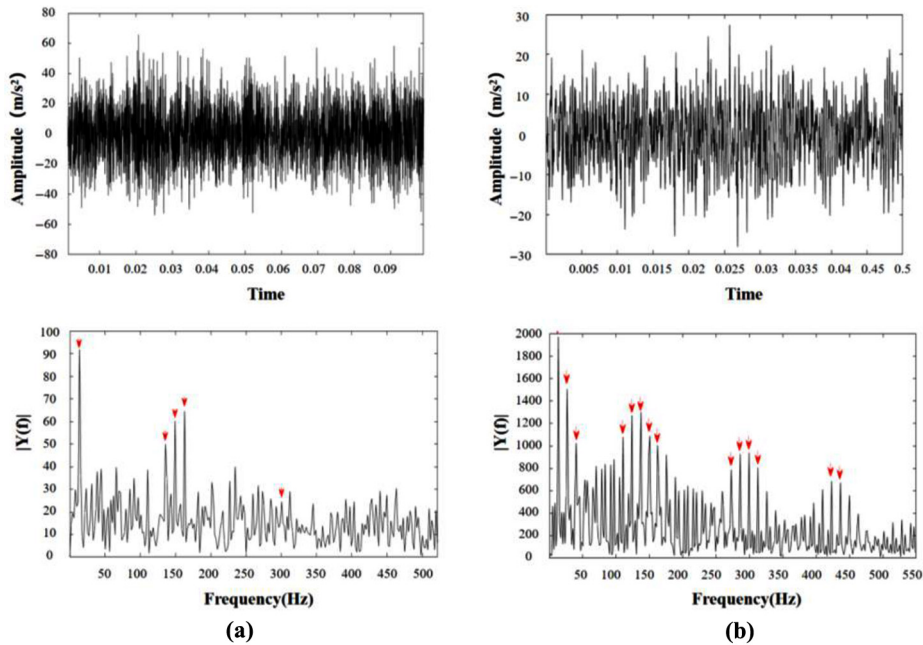


Figure 13. Time domain and frequency domain waveforms of “0.45-inner” type vibration signals in real railway wheelset bearing database

Notes: (a) Signal generated by CGAN; (b) real signal

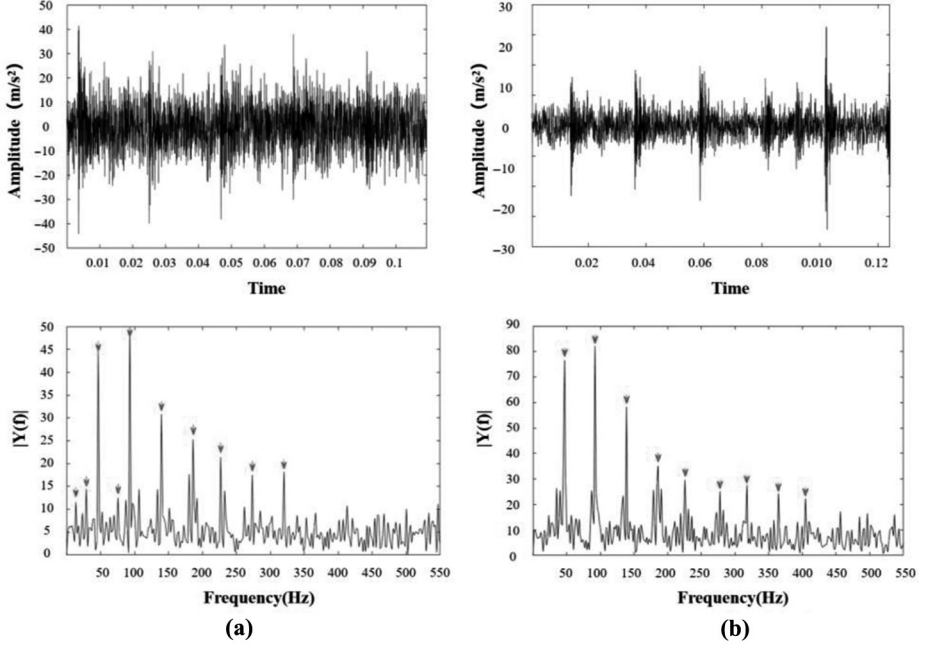


Figure 14.
Time domain and frequency domain waveforms of “0.105-ball” type vibration signals in real railway wheelset bearing database

Notes: (a) Signal generated by CGAN; (b) real signal

on an improved residual network with the channel attention mechanism is constructed. The network structure parameters of the diagnostic model are shown in Table 1. The diagnostic model is implemented by PyTorch. The batch size is 20, the number of training iterations is 2,500 and the initial learning rate is 0.01.

Table 1.
Network structure parameters of the diagnostic model

Serial no.	Layer	Parameters (convolution kernel, stride) × no.
1	Input	–
2	Conv_2D_1	(3*3, 1*1) × 8
3	MaxPooling_1	(2*2, 2*2)
4	SE Block_1	–
5	ResBlock_1_1	(1*1, 1*1) × 8
6	ResBlock_1_2	(3*3, 1*1) × 16
7	ResBlock_1_3	(1*1, 1*1) × 16
8	Conv_2D_2	(3*3, 1*1) × 16
9	Maxpooling_2	(2*2, 2*2)
10	SE Block_2	–
11	ResBlock_2_1	(1*1, 1*1) × 16
12	ResBlock_2_2	(3*3, 1*1) × 32
13	ResBlock_2_3	(1*1, 1*1) × 32
14	Conv_2D_3	(3*3, 1*1) × 32
15	SE Block_3	–
16	FC+Dropout	100, 0.1
17	Output	–

The model was trained and tested using the vibration signal at 1,730 rpm in the CWRU data set. The signal is segmented into nonoverlapping sample segments of length 784 data points, converted into a two-dimensional signal (28×28) and fed into the network. To simulate the imbalanced state of data, the fault samples are artificially reduced. The experimental design has 12 different imbalance scenarios (Table 2). Column 2 in the table gives the proportion of sample imbalance. Columns 3–6 in the table give the number of samples in each class. The last column in the table gives the number of test set samples. The output of the model is a 1×10 vector activated by the softmax function. The elements in the vector correspond to the probabilities that the samples belong to various classes (“0.007-IRF,” “0.007-ORF,” “0.007-BF,” “0.014-IRF,” “0.014-ORF,” “0.014-BF,” “0.021-IRF,” “0.021-ORF,” “0.021-BF” and “Normal”).

To demonstrate the superiority of this method by comparison, the following diagnostic methods were implemented:

- Nosampling CNN: This method does not handle data imbalances. And the diagnosis model removes the residual blocks and SE blocks based on the diagnosis model proposed in this paper.
- Nosampling SE_Res: This method also does not deal with the data imbalance problem. And the diagnostic model uses the model proposed in this paper.
- SMOTE SE_Res: This method handles data unbalances by SMOTE. And the diagnostic model uses the model proposed in this paper.
- ADASYN SE_Res: This method handles data unbalances by ADASYN. And the diagnostic model uses the model proposed in this paper.
- CGAN SE_Res: This method handles data unbalances by CGAN proposed in this paper. And the diagnostic model uses the model proposed in this paper.

In the experiment, training and testing were repeated ten times, and then, the average diagnostic accuracy of the test set was calculated as the evaluation metric of the methods’ diagnostic performance. Table 3 shows the experimental results. To avoid the injustice of the accuracy metrics because of the unbalanced samples and to visualize the performance of several methods more intuitively, a confusion matrix is also drawn (Figure 15). Specifically, analyzing the experimental results, it can be found that:

Serial no.	Proportion	Normal	0.007IRF	0.007ORF	0.007BF	Test data	
			0.014IRF	0.014ORF	0.014BF		
			0.021IRF	0.021ORF	0.021BF		
1	20:1	1,000	50/50/50	1,000/1,000/1,000	1,000/1,000/1,000	500	
2	10:1	1,000	100/100/100	1,000/1,000/1,000	1,000/1,000/1,000	500	
3	5:1	1,000	200/200/200	1,000/1,000/1,000	1,000/1,000/1,000	500	
4	2:1	1,000	500/500/500	1,000/1,000/1,000	1,000/1,000/1,000	500	
5	20:1	1,000	1,000/1,000/1,000	50/50/50	1,000/1,000/1,000	500	
6	10:1	1,000	1,000/1,000/1,000	100/100/100	1,000/1,000/1,000	500	
7	5:1	1,000	1,000/1,000/1,000	200/200/200	1,000/1,000/1,000	500	
8	2:1	1,000	1,000/1,000/1,000	500/500/500	1,000/1,000/1,000	500	
9	20:1	1,000	1,000/1,000/1,000	1,000/1,000/1,000	50/50/50	500	
10	10:1	1,000	1,000/1,000/1,000	1,000/1,000/1,000	100/100/100	500	
11	5:1	1,000	1,000/1,000/1,000	1,000/1,000/1,000	200/200/200	500	
12	2:1	1,000	1,000/1,000/1,000	1,000/1,000/1,000	500/500/500	500	

Table 2.
Details of 12 different
imbalance scenarios
of CWRU data set

Table 3.
Summary of
diagnostic results of
CWRU dataset

Unbalanced categories	Proportion	<i>Nosampling CNN</i>	<i>Nosampling SE_Res</i>	<i>SMOTE SE_Res</i>	<i>ADASYN SE_Res</i>	<i>CGAN SE_Res</i>
0.007IRF	20:1	69.76	70.38	78.70	80.08	81.28
0.014IRF	10:1	70.08	70.92	85.52	86.28	88.58
0.021IRF	5:1	74.64	74.50	92.06	89.96	91.32
	2:1	85.64	88.24	96.72	92.04	99.70
0.007ORF	20:1	60.34	69.76	80.58	79.04	82.00
0.014ORF	10:1	69.46	71.50	86.62	82.42	87.68
0.021ORF	5:1	75.08	85.48	92.42	89.96	95.72
	2:1	84.42	88.96	97.50	94.08	99.02
0.007BF	20:1	69.88	71.60	83.10	82.28	86.92
0.014BF	10:1	73.74	75.64	87.54	86.74	89.84
0.021BF	5:1	77.28	79.86	93.00	91.46	98.54
	2:1	81.84	90.28	97.06	94.20	99.48

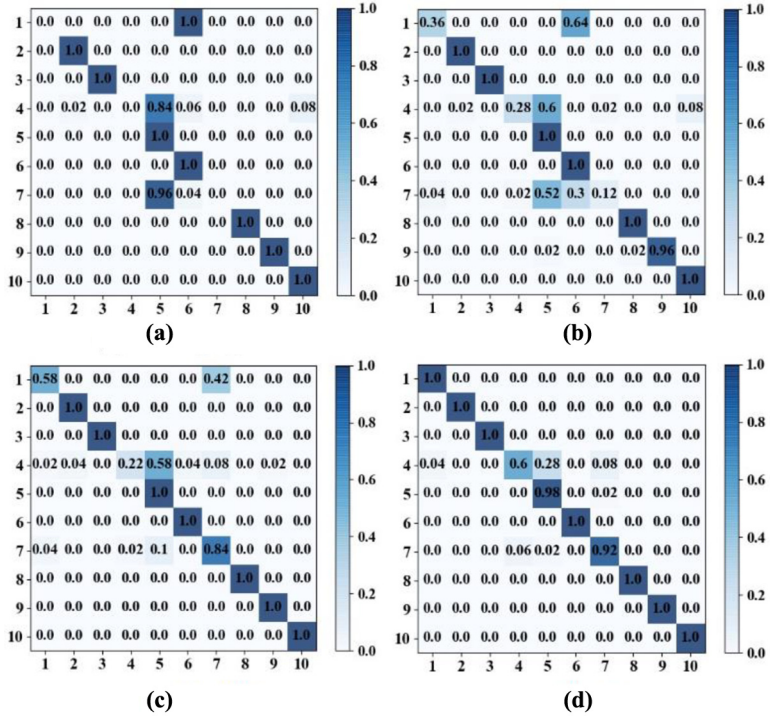


Figure 15.
Confusion matrix of
inner ring data with
imbalanced ratio of
10:1 (CWRU data set)

Notes: (a) No sampling; (b) SMOTE; (c) ADASYN; (d) ours

- By comparing the diagnostic accuracy of *Nosampling SE_Res* and *CGAN SE_Res*, it can be found that if the data imbalance problem is not dealt with, the diagnostic accuracy of the model will be significantly reduced.
- By comparing the diagnostic accuracy of *SMOTE SE_Res*, *ADASYN SE_Res* and *CGAN SE_Res*, it can be indirectly reflected that the quality of data generated by

CGAN is significantly higher than that of traditional data synthesis methods. The reason may be that the data generation mechanism of CGAN is nonlinear and has a stronger data generation ability.

- By comparing the diagnostic accuracy of *Nosampling CNN* and *Nosampling SE_Res*, it can be seen that after introducing SE blocks and residual blocks, the diagnostic performance of the model has been improved. The reason is that the residual block is conducive to gradient propagation to improve the learning effect, while SE Block implements adaptive channel weighting to make the model more flexible, just like the computer version field.
- As the data imbalance problem becomes more severe, the diagnostic performance of *CGAN SE_Res* also deteriorates. It is because more real samples help CGAN to generate higher quality samples. When there are too few real samples, the data generated by CGAN cannot cover all possibilities.

Furthermore, experiments are conducted using the QDSF data set with a lot of noise that is closer to the train operating environment. The data (longitudinal load is 146 kN, the lateral load is 0 and the rotational speed is 120 km/h) are used. Eight data imbalance cases were designed (Table 4). The experimental method is the same as when using the CWRU data set. The only difference is that the output of the model becomes a 1×8 vector because of the change in the data set. The experimental results are shown in Table 5. It can be seen that the experimental results on the QDSF data set also show similar laws to the experimental

Serial no.	Proportion	Normal	0.135IRF 0.45IRF	0.3ORF 0.45ORF	0.001BF 0.105BF	CF	Test data
1	20:1	1,000	1,000/1,000	50/50	1,000/1,000	1,000	500
2	10:1	1,000	1,000/1,000	100/100	1,000/1,000	1,000	500
3	5:1	1,000	1,000/1,000	200/200	1,000/1,000	1,000	500
4	2:1	1,000	1,000/1,000	500/500	1,000/1,000	1,000	500
5	20:1	1,000	1,000/1,000	50/50	1,000/1,000	1,000	500
6	10:1	1,000	1,000/1,000	100/100	1,000/1,000	1,000	500
7	5:1	1,000	1,000/1,000	200/200	1,000/1,000	1,000	500
8	2:1	1,000	1,000/1,000	500/500	1,000/1,000	1,000	500

Table 4.
Details of eight
different imbalance
scenarios of QDSF
data set

Unbalanced categories	Proportion	Nosampling CNN	Nosampling SE_Res	SMOTE SE_Res	ADASYN SE_Res	CGAN SE_Res
0.135IRF	20:1	69.85	73.52	81.23	84.63	85.42
0.45IRF	10:1	71.43	73.44	84.68	85.85	87.46
	5:1	70.04	73.39	88.98	86.00	95.17
	2:1	68.44	74.06	88.49	84.19	98.61
0.3ORF		86.79	87.85	88.51	87.67	99.60
	20:1	72.08	73.40	87.63	85.25	88.12
	10:1	72.71	75.18	88.08	86.82	90.00
0.45ORF	5:1	73.32	79.29	88.75	87.04	99.10
	2:1	90.69	91.31	89.01	87.02	99.16
		92.48	93.72	88.96	87.04	99.87

Table 5.
Summary of
diagnostic results of
QDSF data set

results on the CWRU data set. The proposed method also exhibits excellent diagnostic ability in the case of low signal-to-noise ratio data sets.

In general, the following conclusions can be drawn from the two groups of experiments:

- The CGAN data generation method proposed in this paper outperforms the SMOTE method and the ADASYN method in solving the data imbalance problem.
- The diagnostic model based on residual network with SE module proposed in this paper has higher diagnostic accuracy than the traditional CNN model.
- The combined method of CGAN and the proposed diagnostic model can solve the problem of bearing fault diagnosis under the condition of data imbalance and low signal-to-noise ratio and the effect is excellent.
- Although CGAN can generate high-quality samples, it should still try to alleviate the data imbalance problem during the data collection stage.

6. Conclusion

Aiming at the data imbalance problem based on the fact that the actual bearing fault samples are far less than the normal samples, this paper proposes a solution that uses the CGAN to generate high-quality fault samples to construct a balanced data set. Furthermore, this paper proposes a deep convolutional neural network that fuses residual blocks and SE blocks for bearing fault diagnosis. This model can be used in conjunction with the CGAN generative model, which shows better diagnostic ability in the case of data imbalance and low signal-to-noise ratio compared to other traditional methods. Of course, there are still many problems in the diagnosis of train bearings. For example, laboratory data is often different from the data collected from the train, because the data collected from the real vehicle may be disturbed by vibration of other parts of the train, track irregularities, wheel out-of-roundness or other sources. So how to transfer and apply the diagnostic model trained on laboratory data to the actual train requires more in-depth research. Therefore, in future research, we will pay more attention to this direction with the goal of putting our model into practical application.

References

- Ahsan, R., Shi, W., Ma, X. and Lee Croft, W. (2022), "A comparative analysis of CGAN-based oversampling for anomaly detection", *IET Cyber-Physical Systems: Theory and Applications*, Vol. 7 No. 1, pp. 40-50.
- Antoni, J. (2007), "Fast computation of the kurtogram for the detection of transient faults", *Mechanical Systems and Signal Processing*, Vol. 21 No. 1, pp. 108-124.
- Chen, Z.Y., Gryllias, K. and Li, W.H. (2019), "Mechanical fault diagnosis using convolutional neural networks and extreme learning machine", *Mechanical Systems and Signal Processing*, Vol. 133, p. 106272.
- Chen, R., Zhu, J., Hu, X., Wu, H., Xu, X. and Han, X. (2021), "Fault diagnosis method of rolling bearing based on multiple classifier ensemble of the weighted and balanced distribution adaptation under limited sample imbalance", *ISA Transactions*, Vol. 114, pp. 434-443.
- Gonzalez, S., Garcia, S., L.I., S.-T. and Herrera, F. (2019), "Chain based sampling for monotonic imbalanced classification", *Information Sciences*, Vol. 474, pp. 187-204.
- Goodfellow, I.J., Pouget-Abadie, J., Mirza, M. and Xu, B. (2014), "Generative adversarial networks".
- He, K., Zhang, X., Ren, S. and Sun, J. (2016), Deep residual learning for image recognition. *29th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, June 26, 2016 – July 1, 2016 Las Vegas, NV, United states*. IEEE Computer Society, pp. 770-778.

-
- Hu, J., Shen, L., Albanie, S., Sun, G. and Wu, E. (2020), "Squeeze-and-excitation networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 42 No. 8, pp. 2011-2023.
- Kou, L.L., Qin, Y., Zhao, X.J. and Chen, X.A. (2020), "A multi-dimension end-to-end CNN model for rotating devices fault diagnosis on high-speed train bogie", *IEEE Transactions on Vehicular Technology*, Vol. 69 No. 3, pp. 2513-2524.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K. and Dollar, P. (2020), "Focal loss for dense object detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 42 No. 2, pp. 318-327.
- Liu, F., Chen, R., Xing, K., Ding, S. and Zhang, M. (2022), "Fast fault diagnosis algorithm for rolling bearing based on transfer learning and deep residual network", *Zhendong yu Chongji/Journal of Vibration and Shock*, Vol. 41, pp. 154-164.
- Liu, R.N., Yang, B.Y. and Hauptmann, A.G. (2020), "Simultaneous bearing fault recognition and remaining useful life prediction using joint-loss convolutional neural network", *IEEE Transactions on Industrial Informatics*, Vol. 16 No. 1, pp. 87-96.
- Neupane, D., Kim, Y. and Seok, J. (2021), "Bearing fault detection using scalogram and switchable normalization-based CNN (SN-CNN)", *IEEE Access*, Vol. 9, pp. 88151-88166.
- Niu, Z., Zhong, G. and Yu, H. (2021), "A review on the attention mechanism of deep learning", *Neurocomputing*, Vol. 452, pp. 48-62.
- Qiao, H.H., Wang, T.Y., Wang, P., Zhang, L. and Xu, M.D. (2019), "An adaptive weighted multiscale convolutional neural network for rotating machinery fault diagnosis under variable operating conditions", *IEEE Access*, Vol. 7, pp. 118954-118964.
- Raghuwanshi, B.S. and Shukla, S. (2020), "SMOTE based class-specific extreme learning machine for imbalanced learning", *Knowledge-Based Systems*, Vol. 187, p. 104814.
- Udmale, S.S., Patil, S.S., Phalle, V.M. and Singh, S.K. (2019), "A bearing vibration data analysis based on spectral kurtosis and ConvNet", *Soft Computing*, Vol. 23 No. 19, pp. 9341-9359.
- Xiong, S.C., Zhou, H.D., He, S., Zhang, L.L., Xia, Q., Xuan, J.P. and Shi, T.L. (2020), "A novel end-to-end fault diagnosis approach for rolling bearings by integrating wavelet packet transform into convolutional neural network structures", *Sensors*, Vol. 20 No. 17, p. 4965.
- You, W., Shen, C.Q., Wang, D., Chen, L., Jiang, X.X. and Zhu, Z.K. (2020), "An intelligent deep feature learning method with improved activation functions for machine fault diagnosis", *IEEE Access*, Vol. 8, pp. 1975-1985.
- Zhang, X., Zhao, B. and Lin, Y. (2021), "Machine learning based bearing fault diagnosis using the case western reserve university data: a review", *IEEE Access*, Vol. 9, pp. 155598-155608.
- Zhao, M., Kang, M., Tang, B. and Pecht, M. (2018), "Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes", *IEEE Transactions on Industrial Electronics*, Vol. 65 No. 5, pp. 4290-4300.

Corresponding author

Rui Wang can be contacted at: 276657628@qq.com

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com